GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 29, 2002, 11:38:47 ; Search time 34.66 Seconds
(without alignments)
2073.420 Million cell updates/sec

Title: US-08-485-355B-50
Perfect score: 3374
Sequence: 1 MGDAGVASQRPHNRRGTRNV.....GKIAARVRARRARRRAARAN 647

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3374	100.0	647	19	AAW34535	Helicoverpa armige
2	3353	99.4	647	15	AAR49662	Sequence of Heliot
3	289	8.6	634	19	AAW26785	Nudaurelia beta-li
4	289	8.6	634	19	AAW34537	Nudaurelia beta vi
5	139.5	4.1	1231	20	AAY00219	Enterococcus faeca
6	139.5	4.1	1265	20	AAY00218	Enterococcus faeca
7	127	3.8	2016	22	ABB66424	Drosophila melanog
8	121.5	3.6	2344	22	AAU37120	Staphylococcus aur
9	120.5	3.6	889	20	AAY24913	Human ontherin. H
10	120.5	3.6	889	20	AAY21687	Cadherin-like poly
11	120.5	3.6	979	22	AAE11961	Streptococcus agal
12	120	3.6	932	20	AAY00217	Enterococcus faeca
13	120	3.6	969	20	AAY00216	Enterococcus faeca
14	119.5	3.5	907	6	AAP50073	Epstein-Barr virus
15	119.5	3.5	907	21	AAY68009	Epstein-Barr virus
16	119.5	3.5	1425	22	ABB71908	Drosophila melanog
17	117.5	3.5	26926	22	AAU05396	Human titin (conne
18	116	3.4	641	22	AAE00815	Aspergillus oryzae
19	116	3.4	1004	21	AAB23857	Haemophilus influe
20	115.5	3.4	907	16	AAR80144	EBV gp350/220. Ep
21	113.5	3.4	658	21	AAY68010	Epstein-Barr virus
22	113.5	3.4	781	16	AAW08986	Human parvovirus V
23	113.5	3.4	816	12	AAR14444	A.faecalis penicil
24	112.5	3.3	503	22	AAG92291	C glutamicum prote
25	112	3.3	889	22	ABG23879	Novel human diagno
26	111.5	3.3	583	22	AAU63470	Propionibacterium
27	110.5	3.3	1300	22	AAU33407	Enterococcus faeca
28	110.5	3.3	1300	22	AAU35313	Enterococcus faeca
29	110	3.3	703	21	AAB14781	Aspergillus oryzae
30	109	3.2	703	21	AAB14782	Aspergillus sojae
31	109	3.2	780	20	AAY35502	Amino acid sequenc
32	109	3.2	1302	16	AAR84741	Multidrug resistan
33	109	3.2	1560	10	AAP94145	S. cremoris protei
34	108	3.2	591	20	AAY23337	Talaromyces emerso
35	107.5	3.2	1472	22	ABB59423	Drosophila melanog
36	107.5	3.2	2058	22	AAB97070	Human polypeptide
37	107	3.2	550	17	AAR89501	Penicillin acylase
38	107	3.2	550	17	AAR89502	Penicillin acylase
39	107	3.2	550	17	AAR89504	Penicillin acylase
40	107	3.2	550	17	AAR89499	Penicillin acylase
41	107	3.2	550	17	AAR89500	Penicillin acylase
42	107	3.2	690	21	AAY50835	A. oryzae glutamin
43	107	3.2	690	21	AAY50844	A. oryzae glutamin
44	107	3.2	812	12	AAR10047	abaA gene of Asper
45	107	3.2	2037	22	ABB71928	Drosophila melanog

ALIGNMENTS

RESULT 1
AAW34535
ID AAW34535 standard; Protein; 647 AA.
XX
AC AAW34535;
XX
DT 08-JUN-1998 (first entry)
XX
DE Helicoverpa armigera virus coat protein p71.
XX
KW Vaccine; coat protein; p71; insecticide; Ig-like domain.
XX
OS Helicoverpa armigera virus.
XX
FH Key Location/Qualifiers
FT Domain 281..414
FT /label= Immunoglobulin-like domain
XX
PN WO9746666-A1.
XX
PD 11-DEC-1997.
XX
PF 02-JUN-1997; 97WO-AU00349.
XX
PR 31-MAY-1996; 96AU-0000234.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Gordon KH, Hanzlik TN;
XX
DR WPI; 1998-042175/04.
DR N-PSDB; AAT99117.
XX
PT Modified small RNA viruses and virus-like particles - have altered

```
PT   or substituted Ig-like domains to modify host cell tropism, useful
PT   as insecticides and in medicinal applications
XX
PS   Disclosure; Figure 1; 41pp; English.
XX
CC   The sequence is that of coat protein p71, it has an Ig-like domain
CC   and can be used in the production of virus-like particles (VLP). The
CC   VLPs can be used in vaccines where the Ig-like domain has been altered
CC   so that the VLP presents a surface located antigen which is used to
CC   elicit an immune response in a host organism. They can also be used
CC   for controlling the proliferation of a pest insect and potentially as
CC   medicinal delivery agents for cancer treatment and gene therapy.
XX
SQ   Sequence   647 AA;


  Query Match           100.0%;  Score 3374;  DB 19;  Length 647;
  Best Local Similarity  100.0%;  Pred. No. 4.4e-273;
  Matches  647;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MGDAGVASQRPHNRRGTRNVRVSANTVTVNGRRNQRRRTGRQVSPPDNFTAAAQDLAQSL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mgdagvasqrphnrrgtrnvrvsantvtvngrrnqrrrtgrqvsppdnftaaaqdlaqsl 60

Qy      61 DANTVTFPANISSMPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATESARAVGEYSKIPDG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 dantvtfpanissmpefrnwakgkidldsdsigwyfkyldpagatesaravgeyskipdg 120

Qy     121 LVKFSVDAEIREIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKEMSL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 lvkfsvdaeireiyneecpvvtdvsvpldgrqwslsifsfpmfrtayvavanvenkemsl 180

Qy     181 DVVNDLIEWLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSD 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 dvvndliewlnnladwryvvdseqwinftndttyyvrirvlrptydvpdpteglvrtvsd 240

Qy     241 YRLTYKAITCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSEAYALHTLTFARPSSAAAL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 yrltykaitceanmptlvdqgfwiggqyaltptslpqydvseayalhtltfarpssaaal 300

Qy     301 AFVWAGLPQGGTAPAGTPAWEQASSGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPND 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 afvwaglpqggtapagtpaweqassggyltwrhngttfpagsvsyvlpegfalerydpnd 360

Qy     361 GSWTDFASAGDTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVPPSNAYTNTVFRNTLLET 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gswtdfasagdtvtfrqvavdevvvtnnpagggsaptftvrvppsnaytntvfrntllet 420

Qy     421 RPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQLTPASSFGAV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 rpssrrlelpmppadfgqtvannpkieqsllketlgcylvhskmrnpvfqltpassfgav 480

Qy     481 SFNNPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTP 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 sfnnpgyertrdlpdytgirdsfdqnmstavahfrslshscsivtktyqgwegvtnvntp 540

Qy     541 FGQFAHAGLLKNEEILCLADDLATRLTGVYPATDNFAAAVSAFAANMLSSVLKSEATSSI 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 fgqfahagllkneeilcladdlatrltgvypatdnfaaavsafaanmlssvlkseatssi 600

Qy     601 IKSVGETAVGAAQSGLAKLPGLLMSVPGKIAARVRARRARRRAARAN 647
           |||||||||||||||||||||||||||||||||||||||||||||||
Db     601 iksvgetavgaaqsglaklpgllmsvpgkiaarvrarrarrraaran 647


RESULT   2
AAR49662
ID   AAR49662 standard; Protein; 647 AA.
XX
AC   AAR49662;
XX
DT   12-SEP-1994  (first entry)
XX
DE   Sequence of Heliothis armigera RNA 2 P17.
XX
KW   HaSV; RNA 1; small RNA virus; P17.
XX
OS   Heliothis armigera stunt virus.
XX
PN   WO9404660-A.
XX
PD   03-MAR-1994.
XX
PF   13-AUG-1993;   93WO-AU00411.
XX
PR   14-AUG-1992;   92AU-0004081.
PR   08-JUL-1993;   93US-0089372.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (PACI-) PACIFIC SEEDS PTY LTD.
XX
PI   Christian PD,  Gordon KHJ,  Hanzlik TN;
XX
DR   WPI; 1994-083180/10.
DR   N-PSDB; AAQ58523.
XX
PT   Small RNA virus capable of infecting insect species, e.g.
PT   Heliothis - and transgenic plants contg. viral nucleic acid, for
PT   protection against insect pests
XX
PS   Disclosure; Figure 2; 183pp; English.
XX
CC   The inventors claim a virus comprising a genome hybridisable with
CC   the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are
CC   those given in Figs 1 and 2 of the specification. As isolated
CC   protein or polypeptide prepn. of the proteins or polypeptides
CC   derivable from the virus are also claimed.
CC   H. armigera larvae were raised and viral RNA was extracted. The virus
CC   RNAs were reverse transcribed into cDNA. Clone hr236 contains about
CC   88% or RNA 2. A major translation product of apparent mol. wt. 24,000
CC   is obtd. This protein is derived from a mol. wt. 17,000 reading
CC   frame overlapping the slab of the capsid protein gene. The Mr 24,000
CC   protein (referred to as P17) may have a function in modifying or
CC   manipulating the growth characteristics or cell cycle of
CC   HaSV-infected cells.
XX
SQ   Sequence   647 AA;


  Query Match           99.4%;  Score 3353;  DB 15;  Length 647;
  Best Local Similarity  99.5%;  Pred. No. 2.5e-271;
  Matches  644;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MGDAGVASQRPHNRRGTRNVRVSANTVTVNGRRNQRRRTGRQVSPPDNFTAAAQDLAQSL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mgdagvasqrphnrrgtrnvrvsantvtvngrrnqrrrtgrqvsppdnftaaaqdlaqsl 60

Qy      61 DANTVTFPANISSMPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATESARAVGEYSKIPDG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 dantvtfpanissmpefrnwakgkidldsdsigwyfkyldpagatesaravgeyskipdg 120

Qy     121 LVKFSVDAEIREIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKEMSL 180
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db     121 lvkfsvdaeireiyneecpvvtdvsvpldgrqwslsifsfrrfrtayvavanvenkemsl 180

Qy     181 DVVNDLIEWLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSD 240
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db     181 dvvndliewlnnladwryvvdseqwinftndtryyvrirvlrptydvpdpteglvrtvsd 240

Qy     241 YRLTYKAITCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSEAYALHTLTFARPSSAAAL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 yrltykaitceanmptlvdqgfwiggqyaltptslpqydvseayalhtltfarpssaaal 300

Qy     301 AFVWAGLPQGGTAPAGTPAWEQASSGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPND 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 afvwaglpqggtapagtpaweqassggyltwrhngttfpagsvsyvlpegfalerydpnd 360
```

```
Qy     361 GSWTDFASAGDTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVPPSNAYTNTVFRNTLLET 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gswtdfasagdtvtfrqvavdevvvtnnpagggsaptftvrvppsnaytntvfrntllet 420

Qy     421 RPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQLTPASSFGAV 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 rpssrrlelpmppadfgqtvannpkieqsllketlgcylvhskmrnpvfqltpassfgav 480

Qy     481 SFNNPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTP 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 sfnnpgyertrdlpdytgirdsfdqnmstavahfrslshscsivtktyqgwegvtnvntp 540

Qy     541 FGQFAHAGLLKNEEILCLADDLATRLTGVYPATDNFAAAVSAFAANMLSSVLKSEATSSI 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 fgqfahaglLkneeilcladdlatrltgvypatdnfaaavsafaanmlssvlkseatssi 600

Qy     601 IKSVGETAVGAAQSGLAKLPGLLMSVPGKIAARVRARRARRRAARAN 647
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 iksvgetavgaaqsglaklpgllmsvpgkiaarvrarrarrraaran 647
```

RESULT 3
AAW26785
ID AAW26785 standard; Protein; 634 AA.
XX
AC AAW26785;
XX
DT 22-JUN-1998 (first entry)
XX
DE Nudaurelia beta-like virus capsid protein precursor.
XX
KW NBV; RNA virus; transgenic plant; insect resistance;
KW disease resistance; Nadaurelia cytherea capensis;
KW pine tree emperor moth; capsid protein; vector;
KW virus-like particle.
XX
OS Nadaurelia beta-like virus.
XX
PN AU9724669-A.
XX
PD 04-DEC-1997.
XX
PF 02-JUN-1997; 97AU-0024669.
XX
PR 31-MAY-1996; 96AU-0000233.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (UYRH-) UNIV RHODES.
XX
PI Gordon KH, Hanzlik TN, Hendry DA;
XX
DR WPI; 1998-052736/06.
DR N-PSDB; AAV04471.
XX
PT Nudaurelia beta virus nucleic acid - useful for producing
PT recombinant virus, insect-resistant transgenic plants, etc
XX
PS Example 1; Fig 1; 33pp; English.
XX
CC This polypeptide comprises a 70 kDa capsid protein precursor whose
CC amino acid sequence was deduced from an open reading frame identified
CC in the RNA genome (see AAV04471) of Nadaurelia beta-like virus (NBV).
CC The capsid protein precursor is cleaved at a N/G site into 60.5 and
CC 8 kDa capsid proteins. A claimed infectious recombinant insect virus
CC vector comprises an expressible nucleic acid molecule comprising a
CC nucleotide sequence corresponding to all or an infectious and/or
CC insecticidal portion of the genomic RNA of NBV. Also claimed is: a
CC transgenic plant resistant to insect attack that produces NBV such
CC that insects feeding on the plant are deleteriously affected; and a
CC virus-like particle (VLP) prepared from expression of a nucleic acid
CC molecule comprising a sequence encoding the capsid protein of NBV.
CC The invention provides methods for the control of insect pests (the
CC pine-tree emperor moth Nudaurelia cytherea capensis is mentioned).
XX
SQ Sequence 634 AA;

```
  Query Match             8.6%;  Score 289;  DB 19;  Length 634;
  Best Local Similarity   25.7%;  Pred. No. 2.8e-15;
  Matches  177;  Conservative   72;  Mismatches  266;  Indels  174;  Gaps   39;

Qy      35 QRRRTGRQVSPPDNFTAAAQ--DLAQSLDANTVTFPANISSM-------PE--------- 76
           ||    |: |    | |||      : :| : | | : :         ||
Db      31 qradktrekqpepratraaqtttttstqeagsktsprsrtdyqparwpnpeprehpgqpr 90

Qy      77 --FRNWAKGKIDLD---SDSIGWYFKYLDPAGATESARAVGEYSKIPDGLVKFSVDAEIR 131
             |  ||   | :   ||   |   |||| |  :::   |   ||||| :  |   : |
Db      91 sdtregakasddgeshgsdikawihdyldpdgeyktslddg---kipdgaipqstcgqfr 147

Qy     132 EIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKEMSLDVVN-DL---- 186
                | :   ::|||| | | :   | ||   : :    | |  ::| | ||
Db     148 gtvgarypglnsttlpldggtwpllvmhlpffrhpllfitttsnte--vevtnadldafa 205

Qy     187 IEWLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRPT---YDVPDPTE-GLVRTVSDYR 242
            :| ||  ||        |    |   :|: :   ||    ||| ||: |:   :  ||
Db     206 ndw-nnrtdwteat-ypswaqvgn--vfymvv----ptealtdvppptqlgvsgllesyr 257

Qy     243 LTYKAITCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSEAYALHTLT-------FARPS 295
           ||   :|   | ||||:||  :  |:   |    | :  :  |  | |        : |:
Db     258 ltssgvtayfnaptlvnqgvaviaqf--qpdkehqkenpdivagttqtggtlqlggsgpn 315

Qy     296 SAAALAFVWAGLPQGGTA-PAGTPAWEQASSGGYLTWRHNGTTFPAG---SVSYVLPEGF 351
             :  :   :  || | |  | :       | | ::    ||  |   :::  || |
Db     316 ytltmt-igdqvefggaaiplptvsmgpmpesgqlvfqtanltfdvgntititttlppgs 374

Qy     352 ALERYDPNDGSWTDFASAG-DTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVPPSNAYTN 410
                 | |   || | ||||     ||        ||       ||||
Db     375 v-------tgmwqftasngtdtvt-----vd--------aga------tvrv-------- 400

Qy     411 TVFRNTL--LETRPSSRRLEL----PMPPADFGQTVANNPKIEQSLLKETLGCYLVHSKM 464
              |:    | |  :  :|       | | | |    |   |:  | |   | |:  : :
Db     401 ---rsefgrlgtesaghqlnqdstndmnpndag----naktiqfqltkr--ghympeasi 451

Qy     465 RNPVFQLTPASSFGAVSFNN-------------PGYER--TRD-LPDYTGIRDSFDQNM 507
           |    :   |:|:| |  :                || |   || ||  ||        |
Db     452 r---vrndnatsygpvdedtedncgrlhraigalqgyhrqqlrdrlpsmtg--------m 500

Qy     508 STAVAHFRSLSHSCSIVTKTYQGWEGV------TNVNTPFGQFAHAGLLKNEEILCLADD 561
           ||:   :                :||      |    |:| || |   |::  | :|
Db     501 ststvpy---------------lqgvptlrsdtgggepwgpfasatppkddvaltvart 544

Qy     562 LATRLTGVYPATDNFAAAVSAFAANMLSSVLKSEATSSIIKSVGETAVGAAQSGLAKLPG 621
                 ||   |   |: |  |  :: : :       ::|    | | | :    :
Db     545 wtdlhpfayperyngfgalfamvaktiaqipr------yvrs----aagvanavtdcies 594

Qy     622 LLMSVPGKIAARVRARRARR-----RAAR 645
           ||     :  | |||||     | ||
Db     595 atesvasnstserrqrrarrvggiargar 623

RESULT 4
AAW34537
ID AAW34537 standard; Protein; 634 AA.
XX
AC AAW34537;
XX
DT 08-JUN-1998 (first entry)
XX
DE Nudaurelia beta virus coat protein p70.
XX
KW Vaccine; coat protein; p70; insecticide; Ig-like domain.
XX
OS Nudaurelia beta virus.
XX
FH Key Location/Qualifiers
```

```
FT   Domain          285..433
FT                   /note= "Immunoglobulin-like domain"
XX
PN   WO9746666-A1.
XX
PD   11-DEC-1997.
XX
PF   02-JUN-1997;   97WO-AU00349.
XX
PR   31-MAY-1996;   96AU-0000234.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Gordon KH,  Hanzlik TN;
XX
DR   WPI; 1998-042175/04.
DR   N-PSDB; AAT99118.
XX
PT   Modified small RNA viruses and virus-like particles - have altered
PT   or substituted Ig-like domains to modify host cell tropism, useful
PT   as insecticides and in medicinal applications
XX
PS   Disclosure; Figure 2; 41pp; English.
XX
CC   The sequence is that of coat protein p70, it includes an Ig-like domain
CC   which can be used in the production of virus-like particles (VLPs). The
CC   VLPs can be used in vaccines where the Ig-like domain has been altered
CC   so that the VLP presents a surface located antigen which is used to
CC   elicit an immune response in a host organism. They can also be used
CC   for controlling the proliferation of a pest insect and potentially as
CC   medicinal delivery agents for cancer treatment and gene therapy.
XX
SQ   Sequence   634 AA;
```

```
  Query Match             8.6%;  Score 289;  DB 19;  Length 634;
  Best Local Similarity   25.7%;  Pred. No. 2.8e-15;
  Matches  177;  Conservative   72;  Mismatches  266;  Indels  174;  Gaps   39;

Qy      35 QRRRTGRQVSPPDNFTAAAQ--DLAQSLDANTVTFPANISSM-------PE--------- 76
           ||    |:  |    | |||      : :| : | | : :         ||
Db      31 qradktrekqpepratraaqttttttstqeagsktsprsrtdyqparwpnpeprehpgqpr 90

Qy      77 --FRNWAKGKIDLD---SDSIGWYFKYLDPAGATESARAVGEYSKIPDGLVKFSVDAEIR 131
             |  ||   | :   ||   |   |||| |  :::   |   ||||| :  |   : |
Db      91 sdtregakasddgeshgsdikawihdyldpdgeyktsldg---kipdgaipqstcgqfr 147

Qy     132 EIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKEMSLDVVN-DL---- 186
                | :   ::|||| | | :   | ||   : :    | |  ::| | ||
Db     148 gtvgarypglnsttlpldggtwpllvmhlpffrhpllfitttsnte--vevtnadldafa 205

Qy     187 IEWLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRPT---YDVPDPTE-GLVRTVSDYR 242
            :| ||  ||        |    |   :|: :    ||    ||| ||: |:   :  ||
Db     206 ndw-nnrtdwteat-ypswaqvgn--vfymvv----ptealtdvppptqlgvsgllesyr 257

Qy     243 LTYKAITCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSEAYALHTLT-------FARPS 295
           ||   :|   | ||||:||  :  |:   |    | :  :  |  | |         : |:
Db     258 ltssgvtayfnaptlvnqgvaviaqf--qpdkehqkenpdivagttqtggtlqlggsgpn 315

Qy     296 SAAALAFVWAGLPQGGTA-PAGTPAWEQASSGGYLTWRHNGTTFPAG---SVSYVLPEGF 351
             :  :   :  || | |  | :       | | ::    ||  |   :::  || |
Db     316 ytltmt-igdqvefggaaiplptvsmgpmpesgqlvfqtanltfdvgntititttllppgs 374

Qy     352 ALERYDPNDGSWTDFASAG-DTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVPPSNAYTN 410
                 | |   || | ||||     ||        ||       ||||
Db     375 v-------tgmwqftasngtdtvt-----vd--------aga------tvrv-------- 400

Qy     411 TVFRNTL--LETRPSSRRLEL----PMPPADFGQTVANNPKIEQSLLKETLGCYLVHSKM 464
             |:    | |  :  :|       | | | |    |   |:  | |   | |:  : :
Db     401 ---rsefgrlgtesaghqlnqdstndmnpndag----naktiqfqltkr--ghympeasi 451

Qy     465 RNPVFQLTPASSFGAVSFNN-------------PGYER--TRD-LPDYTGIRDSFDQNM 507
           |    :   |:|:| |  :               || |   || ||  ||        |
Db     452 r---vrndnatsygpvdedtedncgrlhraigalqgyhrqqlrdrlpsmtg--------m 500

Qy     508 STAVAHFRSLSHSCSIVTKTYQGWEGV------TNVNTPFGQFAHAGLLKNEEILCLADD 561
           ||:   :                 :||      |    |:| || |   |::  | :|
Db     501 ststvpy---------------lqgvptlrsdtgggepwgpfasatppkddvaltvart 544

Qy     562 LATRLTGVYPATDNFAAAVSAFAANMLSSVLKSEATSSIIKSVGETAVGAAQSGLAKLPG 621
                 ||   |   |: | |  :: : :       ::|    | | | :    :
Db     545 wtdlhpfayperyngfgalfamvaktiaqipr------yvrs----aagvanavtdcies 594

Qy     622 LLMSVPGKIAARVRARRARR-----RAAR 645
             ||     :  | |||||     | ||
Db     595 atesvasnstserrqrrarrvggiargar 623
```

```
RESULT   5
AAY00219
ID   AAY00219 standard; Protein; 1231 AA.
XX
AC   AAY00219;
XX
DT   20-APR-1999  (first entry)
XX
DE   Enterococcus faecalis antigenic polypeptide fragment EF108.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C,  Choi GH,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   N-PSDB; AAX20209.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 9; Page 211; 301pp; English.
XX
CC   The present sequence represents an antigenic polypeptide fragment
CC   isolated from Enterococcus faecalis. The present invention describes
CC   genes, proteins and antigenic polypeptides isolated from E. faecalis.
CC   The proteins can be used in vaccines for preventing or attenuating an
CC   infection caused by a member of the Enterococcus genus in an animal.
CC   They can also be used for detecting Enterococcus antibodies in a sample.
CC   The nucleotide sequences can be used for detecting Enterococcus nucleic
CC   acids. Products from the present invention can also be used for
CC   screening compounds to identify agonists and antagonists of E. faecalis
CC   protein activity.
XX
SQ   Sequence   1231 AA;
```

```
  Query Match             4.1%;  Score 139.5;  DB 20;  Length 1231;
  Best Local Similarity   19.0%;  Pred. No. 0.024;
  Matches  141;  Conservative   94;  Mismatches  252;  Indels  257;  Gaps   31;

Qy      19 NVRVSANTVTVNGRRNQRRRTGRQVSPPDNFTAAAQDLAQSLDANTV----TFPANISS- 73
           | :|: | |  || :         ::||  ||   : :  | || |     | |   ::
```

```
Db    542 nrkvtenfvdtngak---------itpptgftqgkktvits-daytfkqagtlpdtyttg 591

Qy     74 --MPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATESARAVGEYSKIPDGLVKFSVDAEIR 131
          :|: | |||  |:: :                       :| |   | :  : ::
Db    592 gktykfkgwykgksilntltt---------------------tkapsyqvtyddnddln 629

Qy    132 EIYNEECPVVTDVSVPLD---------------GRQWSLSIFSFPMFRTAYVAVANVEN 175
          :| ||      || ::               |: :: |  ::   ||    | :  |
Db    630 vvyeeetvttvypsvdmnfvnekggaftpaltfsgkyyaqstsay--lrtdlydvtsknn 687

Qy    176 ---------KEMSLDVVNDLIEWLN---------------------------------- 191
                    |: :  :|::  |
Db    688 gngqytvsinngsmplsqellkkynngqpisatnrlqfnvdklaidqqlkyvdsiqldta 747

Qy    192 ---NLADWRYVV--------------------DSEQWINFTNDTTYYV----RIRVLRP 223
          ||  :|||                      :|   :|| :| ||:     |:
Db    748 qssnlksyryvytnnsslvfdpnvapaevdlsseslnllnfdsdgtyfsnannrlfythl 807

Qy    224 TYD-VPDPTEGLVRTVSD----------YRLTYKAITCEANMPTLVDQGFWIGGQYALTP 272
          |   |     ||  : :          |::| | :|        ||    : |     |
Db    808 gysgtpgvnyllvmflfnakpadkskl vykvtrkqvt-----enfvd----vngakitap 858

Qy    273 TSLPQYDVSEAYALHTLTFARPSSAAAL---------AFVWAGLPQGGTAPAGTPAWEQA 323
          |  ||        ::: || : ::| ||          : : |  :| | |:
Db    859 tgftq---gnqvpmnsntf-kytaakalpatyttggkvytfqgwykgktkpstl------ 908

Qy    324 SSGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFRQVAVDEV 383
                | || |  : :           :| ||     :      | :        ||
Db    909 ----------nktttptfnat-----------fdgnddmtamykeeiptasvtltrpkev 947

Qy    384 VVTNNPAGGGSAPTFTVRVPPSNAYTNTVFRNTLLETRPS-SRRLELP--MPPADFGQTV 440
          : ||      :  | | : |          :|  |:  |: |  | :|  |     |:|
Db    948 idtntnviwtttitntskap---------lqnltlkkgpnwsagltiptfmevtpegett 998

Qy    441 ANNPKIEQSLLKETLGCYLVHSKMRNPVFQLTPASSFGAVSF-----NNPGYERTRDLPD 495
          : | :  :|  | :           |:    |     :|:|       |      ::
Db    999 ksip-vnstlwtegv----------plpnavpigkkvsvafttratgkpntvlkaevvv 1046

Qy    496 YTGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEI 555
          : ||:||   |       |   :  :|| | :|:  |   :  |||   ||  :   :
Db   1047 fggikdstvdnf------vrirpndqevvtpttegfisvptfd--fgqvgvagtkqqhsl 1098

Qy    556 LCLAD--------------------DLATRLTGVYPATDNF---------AAAVSAFAAN 586
          ||                       |  :|:    |||:           || ||:|
Db   1099 kqaadyygngtrnpylrikktqpnwsltaqlsqpksatdslptatrlllgaapvssftny 1158

Qy    587 MLSSVLKSE--ATSSIIKSVGETA 608
          : ||:    ||:|  :   ||
Db   1159 nqptelkntvgttsaisltannta 1182


RESULT   6
AAY00218
ID   AAY00218 standard; Protein; 1265 AA.
XX
AC   AAY00218;
XX
DT   20-APR-1999  (first entry)
XX
DE   Enterococcus faecalis protein EF108.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C,  Choi GH,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   N-PSDB; AAX20208.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 9; Page 209-210; 301pp; English.
XX
CC   The present sequence represents a protein isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
XX
SQ   Sequence   1265 AA;


  Query Match             4.1%;  Score 139.5;  DB 20;  Length 1265;
  Best Local Similarity   19.0%;  Pred. No. 0.025;
  Matches  141;  Conservative   94;  Mismatches  252;  Indels  257;  Gaps   31;

Qy     19 NVRVSANTVTVNGRRNQRRRTGRQVSPPDNFTAAAQDLAQSLDANTV----TFPANISS- 73
          | :|: | |  || :          ::||  ||   : :  | || |     | |   ::
Db    574 nrkvtenfvdtngak---------itpptgftqgkktvits-daytfkqagtlpdtyttg 623

Qy     74 --MPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATESARAVGEYSKIPDGLVKFSVDAEIR 131
          :|: | |||  |:: :                       :| |   | :  : ::
Db    624 gktykfkgwykgksilntltt---------------------tkapsyqvtyddnddln 661

Qy    132 EIYNEECPVVTDVSVPLD---------------GRQWSLSIFSFPMFRTAYVAVANVEN 175
          :| ||      || ::               |: :: |  ::   ||    | :  |
Db    662 vvyeeetvttvypsvdmnfvnekggaftpaltfsgkyyaqstsay--lrtdlydvtsknn 719

Qy    176 ---------KEMSLDVVNDLIEWLN---------------------------------- 191
                    |: :  :|::  |
Db    720 gngqytvsinngsmplsqellkkynngqpisatnrlqfnvdklaidqqlkyvdsiqldta 779

Qy    192 ---NLADWRYVV--------------------DSEQWINFTNDTTYYV----RIRVLRP 223
          ||  :|||                      :|   :|| :| ||:     |:
Db    780 qssnlksyryvytnnsslvfdpnvapaevdlsseslnllnfdsdgtyfsnannrlfythl 839

Qy    224 TYD-VPDPTEGLVRTVSD----------YRLTYKAITCEANMPTLVDQGFWIGGQYALTP 272
          |   |     ||  : :          |::| | :|        ||    : |     |
Db    840 gysgtpgvnyllvmflfnakpadkskl vykvtrkqvt-----enfvd----vngakitap 890

Qy    273 TSLPQYDVSEAYALHTLTFARPSSAAAL---------AFVWAGLPQGGTAPAGTPAWEQA 323
          |  ||        ::: || : ::| ||          : : |  :| | |:
Db    891 tgftq---gnqvpmnsntf-kytaakalpatyttggkvytfqgwykgktkpstl------ 940

Qy    324 SSGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFRQVAVDEV 383
                | || |  : :           :| ||     :      | :        ||
Db    941 ----------nktttptfnat-----------fdgnddmtamykeeiptasvtltrpkev 979

Qy    384 VVTNNPAGGGSAPTFTVRVPPSNAYTNTVFRNTLLETRPS-SRRLELP--MPPADFGQTV 440
          : ||      :  | | : |          :|  |:  |: |  | :|  |     |:|
Db    980 idtntnviwtttitntskap---------lqnltlkkgpnwsagltiptfmevtpegett 1030

Qy    441 ANNPKIEQSLLKETLGCYLVHSKMRNPVFQLTPASSFGAVSF-----NNPGYERTRDLPD 495
          : | :  :|  | :           |:    |     :|:|       |      ::
Db   1031 ksip-vnstlwtegv----------plpnavpigkkvsvafttratgkpntvlkaevvv 1078
```

```
Qy    496 YTGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEI 555
          : ||:||   |       |   :   :|| | :|:  |   :  |||   ||  :   :
Db   1079 fggikdstvdnf------vrirpndqevvtpttegfisvptfd--fgqvgvagtkqqhsl 1130

Qy    556 LCLAD-------------------DLATRLTGVYPATDNF---------AAAVSAFAAN 586
            ||                    |  :|:    |||:          || ||:|
Db   1131 kqaadyygngtrnpylrikktqpnwsltaqlsqpksatdslptatrlllgaapvssftny 1190

Qy    587 MLSSVLKSE--ATSSIIKSVGETA 608
          : ||:    ||:|  :   ||
Db   1191 nqptelkntvgttsaisltannta 1214


RESULT   7
ABB66424
ID   ABB66424 standard; Protein; 2016 AA.
XX
AC   ABB66424;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 26064.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10527.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 26064; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2016 AA;


  Query Match             3.8%;  Score 127;  DB 22;  Length 2016;
  Best Local Similarity   20.7%;  Pred. No. 0.57;
  Matches   77;  Conservative   45;  Mismatches  116;  Indels  134;  Gaps   18;

Qy    231 TEGLVRTVSDYRLTYKAITCEANMPTLVDQGFWI-------GGQYALTPTSLPQYDVSEA 283
          ||   : |  |:::::|   | | |      ||:        || : :  ::|   |
Db   1259 tetktqkvphyqmsfeateleknkp----yefwvtasttigegqqsksivampsdqvpak 1314

Qy    284 YA----LHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLTWRHNGTTFP 339
           |      | ||   :    ||    | ||                    :||:  |  |
Db   1315 iasfddtftatfkedakmpcla---vgapq----------------peitwkikgvefs 1354

Qy    340 AGSVSYVLPEGFAL----ERYDPNDGS-WTDFASAGDTVTFRQVAV------------- 380
          |     |||:|  |     | |  | |   : : | |::| : : :
Db   1355 andrmrvlpdgslliksvnrqdagdyschaensiakdsithklivlappqsphvtlsatt 1414

Qy    381 -DEVVVTNNPAGGGSAPT--FTVRVPPS--------------------------NAYTN 410
           | : |   |  | :||   :|:   |                           |
Db   1415 tdaltvklkphegdtaplhgytlhykpefgewetsevsvdsqkhniegllcgsryqvya- 1473

Qy    411 TVFRN-------TLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGC----YL 459
          | | |        :| ||   :: :||  |  | :  :|:  :     |:  ||    ::
Db   1474 tgfnnigageasdilntrtkgqkpklpekpr-fievssnsvslhfkawkdg-gcpmshfv 1531

Qy    460 VHSKMRNP-----------------VFQLTPASSFG--AVSFNNPGY------------- 487
          | || |:                  |  | ||: :     : |: |:
Db   1532 veskkrdqiewnqisnnvkpdnnyvvldlepatwynlritahnsagftvaeydfatltvt 1591

Qy    488 ----ERTRDLPD 495
               :||||:
Db   1592 ggtiapsrdlpe 1603


RESULT   8
AAU37120
ID   AAU37120 standard; Protein; 2344 AA.
XX
AC   AAU37120;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1290.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54979.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12713; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
```

```
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2344 AA;


  Query Match             3.6%;  Score 121.5;  DB 22;  Length 2344;
  Best Local Similarity   20.2%;  Pred. No. 2.1;
  Matches  142;  Conservative   81;  Mismatches  322;  Indels  159;  Gaps   31;

Qy      22 VSANTVTVNGRRNQRRRTGRQVSPPDNFTAA---AQDLAQSLDANTVTFPANISSMPEFR 78
           ::|||:|||    ::  |    :  |  |      ||       |  :|    | |   |
Db     254 ltantitvnkdnlkqymttsgnatydqstgvvtltqdttsqkga--itlgtridsnksfh 311

Qy      79 NWAKGKIDLDS---------DSIGWYFK--YLDPAGATESARAVGEYSKIPDGLVKFSVD 127
              ||::| :         | ||: |    |   |   :|  :|  |         | :|
Db     312 --fsgkvnlgnkyeghgnggdgigfafspgvlgetglngaavgiggls----nafgfkld 365

Qy     128 AEIREIYNEECP-----VVTDVSVPLDGRQWSLSIFSFPMFRT--------AYVAVANVE 174
               :|   |        | |    |        :|  | |         | : :  :
Db     366 t----yhntstpnssakakadpsnvaggg------afgafvttdsygvastytssstad 414

Qy     175 N-KEMSLDVVNDLIE--------------------WLNNLADWRYVVDSEQWINF---- 208
           |  ::::   |:  :                    |  |::||   :      ||
Db     415 naaklnvqptnntfqdfdinyngdtkvmtvkyagqtwtrnisdw---iaksgttnfslsm 471

Qy     209 ---TNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCEANMPTLVDQGFWIG 265
              |   |   :::    |::  :     || |     | | |         |||   |
Db     472 tastggatnlqqvqf--gtfeytesavtqvryvd--vttgkdiippktysgnvdqvvtid 527

Qy     266 GQY-ALTPTSLPQYDVSEAYAL------HTLTFARPSSAAALAFVWAGLPQGGTAPAGTP 318
           |  |||        |  :||        |:       :    |     |   |   |
Db     528 nqqsaltakgynytsvdssyastyndtnktvkmtnaggsvtyyftdvkap---tvtvgnq 584

Qy     319 AWEQASSGG--YLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGS-------------- 362
             |   :     ||   |||     :|:  || | :   ||    |
Db     585 tievgktmnpivltttdngtgtvtntvt-glpsgls---ydsatnsiigtptkigqstvt 640

Qy     363 --WTDFASAGDTVTFRQVAVDEVVVTNNPAGGGSAPTFTVRVP--------PSNAYTNTV 412
              || |:   | ||    ||     |  | |  |:  |:   |          || ||||
Db     641 vvstdqannksttftfinvvdttaptvtpigdkssevfspispiniatqdnsgnavtntv 700

Qy     413 FRNTLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQLT 472
                 ||    :          |::  |                 : :      :
Db     701 ------tglpsgltfd------stnntisgtp----------------tnigtstitiv 731

Qy     473 PASSFGAVSFNNPGYERTRD-LPDYTGIRDSFDQNMST----AVAHFRSLSHSCSIVTKT 527
             :|  :     || ||: : |      |  |: |     | :   | | | ||:| |
Db     732 stdasgnkttttfkyevtrnsmsdsvstsgstqqsqsvstskadsqsaststsgsimtst 791

Qy     528 YQGWEGVTNVNTPFGQFAHAGLLKNEEILCLADDLATRLTGVYPATDNFAAAVSAFAANM 587
                |:|:      |    |   |   ::   :| |      : : : :||   :
Db     792 sastskstsvslsdsvsasks-lstsesnsvsssststslvnsqsvsssmsgsvs--ksts 848

Qy     588 LSSVLKSEATSSIIKSVGETAVGAAQSGLAKLPGLLMSVPGKIA 631
           ||  : : :::   :||  :   : ::  :    : ||  | ::
Db     849 lsdfisnsssteksesvststsdslrtstslsdsvsmstsgsls 892


RESULT   9
AAY24913
```

```
ID   AAY24913 standard; Protein; 889 AA.
XX
AC   AAY24913;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Human ontherin.
XX
KW   Human; ontherin; cadherin-related gene; cadherin-like protein;
KW   tissue homeostasis; neurological condition; nervous system; ischaemia;
KW   infection; inflammation; Alzheimer's disease; neurodegenerative disease;
KW   Parkinson's disease; Huntington's chorea; amylotrophic lateral sclerosis;
KW   spinocerebellar degeneration; immunological disease; multiple sclerosis.
XX
OS   Homo sapiens.
XX
PN   WO9929860-A1.
XX
PD   17-JUN-1999.
XX
PF   08-MAY-1998;   98WO-US09151.
XX
PR   08-DEC-1997;   97US-0067887.
XX
PA   (ONTO-) ONTOGENY INC.
XX
PI   Israel D;
XX
DR   WPI; 1999-385606/32.
DR   N-PSDB; AAX83610.
XX
PT   New isolated cadherin-like polypeptide, ontherin
XX
PS   Claim 20; Page 83-86; 91pp; English.
XX
CC   The present sequence represents human ontherin (OT), which is a
CC   cadherin-like protein. The OT protein can: (i) bind to Ca2+, a
CC   catenin or a cadherin; (ii) regulate differentiation of neuronal cells;
CC   (iii) regulates survival of differentiated neuronal cells; (iv)
CC   regulates proliferation of testicular germ line cells; and/or (v)
CC   regulates proliferation of renal cells. The OT protein has vasotropic,
CC   cytostatic, anti-inflammatory, nootropic and neuroprotective activity.
CC   The OT protein is involved in the formation and maintenance of ordered
CC   spatial arrangements of differentiated tissues in vertebrates, both
CC   adult and embryonic, and can be used to generate and/or maintain an
CC   array of different vertebrate tissue both in vitro and in vivo. OT
CC   therapeutics can be used for treating e.g. neurological conditions such
CC   as: (i) acute, subacute, or chronic injury to the nervous system,
CC   including traumatic injury, chemical injury, vasal injury and deficits
CC   (such as the ischaemia resulting from stroke), together with
CC   infectious/inflammatory and tumour induced injury; (ii) aging of the
CC   nervous system including Alzheimer's disease; (iii) chronic
CC   neurodegenerative diseases of the nervous system including Parkinson's
CC   disease, Huntington's chorea, amylotrophic lateral sclerosis, as well as
CC   spinocerebellar degenerations; and (iv) chronic immunological diseases
CC   of the nervous system or affecting the nervous system including multiple
CC   sclerosis, for selective ablation of sensory neurons, e.g. in the
CC   treatment of chronic pain syndromes, or in the treatment of neoplastic
CC   or hyperplastic transformations such as may occur in the central nervous
CC   system.
XX
SQ   Sequence   889 AA;


  Query Match             3.6%;  Score 120.5;  DB 20;  Length 889;
  Best Local Similarity   20.4%;  Pred. No. 0.57;
  Matches  144;  Conservative   92;  Mismatches  250;  Indels  219;  Gaps   36;

Qy      32 RRNQRRRTGRQVSPPDNFTAAAQDLAQSLDANTVTFPANISSMPEFRNWAKG-------- 83
           | | : :   :|   |       :   | :: |  :| ::   :    | | |
Db      97 rhnakcqlslevfandkeicmikveiqdindnapsfssdqidldisenaapgtrfpltsa 156

Qy      84 -KIDLDSDSIGWYFKYLDPAGATE-SARAVGEYSKIPDGLVKFSVDAEIREIYNEECPVV 141
```

```
Db    157 hdpdagenglrtylltrddhglfgldvksrgdgtkfpelviqkald---reqqnhhtlvl 213

Qy    142 TDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKEMS--LDVVNDLIEWLNNLADWRYV 199

Db    214 t----aldgge-------pprsatvqinvkvidsndnspvfeapsylvelpentplgtvv 262

Qy    200 VDSEQWINFTN-------DTTY----YVRIRVLRPTYDVPDPTEGLVRTVS--DY----- 241

Db    263 id----lnatdadegpngevlysfssyvpdrv-relfsi-dpktglirvkgnldyeengm 316

Qy    242 ------------------RLTYKAITCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSE 282

Db    317 leidvqardlgpnlipahckvtvklidrndnapsi---gf----------vsvrqgalse 363

Qy    283 AYALHTLTFARPSSAAALAFVW---AG---------LPQGGTAPAGTPAWEQASSGGYLT 330

Db    364 a--------appgtvialvrvtdrdsgkngqlqcrvlggggtgggg-------------- 401

Qy    331 WRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFRQV---AVDEVVVTN 387

Db    402 ----glggpggsvpfkleeny--------dnfyt-------vvtdrpmdretqdeynvti 442

Qy    388 NPAGGGSAP-----TFTVRV------PP---SNAYTNTVFRNTLLETRPSSRRLELPMPP 433

Db    443 vardggspplnstksfaikilddndnpprftkglyvlqvhenni--------------pg 488

Qy    434 ADFGQTVANNPKIEQ------SLLKETLGCYLVHSKMR-NPVFQLTPASSFGAVSFNNPG 486

Db    489 dslgsvlaqdpdlgqngtvsysilpshigdvsiytyvsvnp----tngaiyalrsfn--- 541

Qy    487 YERTRDLPDYTGIRDS-----FDQNMSTAVAHFRSLSHSCSIVTKTYQGWEG-------- 533

Db    542 feqtkafefkvlakdsgapahlesnatvrvtvldvndnapvivlptlqndtaelqvprna 601

Qy    534 -----VTNVNTPFGQFAHAGLLK------NEEILCLADDLATRLTGVYPATDN------F 576

Db    602 glgylvstvraldsdfgesgrltyeivdgnddhlfeidpssgeirtlhpfwedvtpvvel 661

Qy    577 AAAVSAFAANMLSSVLKSEATSSIIKSVGETAVGAAQSGLAKLPG 621

Db    662 vvkvtdhgkptlsavak-----liirsvs----gslpegvprvng 697
```

```
RESULT 10
AAY21687
ID   AAY21687 standard; Protein; 889 AA.
XX
AC   AAY21687;
XX
DT   18-AUG-1999  (first entry)
XX
DE   Cadherin-like polypeptide, ontherin.
XX
KW   Ontherin; cadherin-like polypeptide; cadherin; cell differentiation;
KW   neuronal cell; testicular; renal; spermatogenesis; vertebrate; stroke;
KW   nervous system; neurological; injury; ischemia; inflammatory; tumour;
KW   Alzheimer's disease; neurodegenerative disease; Parkinson's disease;
KW   Huntington's chorea; amylotrophic lateral sclerosis; multiple sclerosis;
KW   spinocerebellar degeneration; pain syndrome; drug screening.
XX
OS   Vertebrata.
XX
PN   WO9929853-A1.
XX
PD   17-JUN-1999.
XX
PF   08-DEC-1998;   98WO-US25981.
XX
PR   08-DEC-1997;   97US-0067887.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Israel DI;
XX
DR   WPI; 1999-385603/32.
DR   N-PSDB; AAX77400.
XX
PT   New isolated cadherin-like polypeptides useful for treating
PT   Alzheimer's disease
XX
PS   Claim 20; Page 96-101; 108pp; English.
XX
CC   This represents a cadherin-like polypeptide, ontherin, which can bind
CC   to at least one of Ca2+, a catenin or a cadherin. The ontherin (OT)
CC   protein regulate differentiation of neuronal cells; regulate survival of
CC   differentiated neuronal cells; regulate proliferation of testicular germ
CC   line cells; and/or regulate proliferation of renal cells. The polypeptide
CC   preferably regulate spermatogenesis. The OT polypeptides are involved in
CC   the formation and maintenance of ordered spatial arrangements of
CC   differentiated tissues in vertebrates, both adult and embryonic, and can
CC   be used to generate and/or maintain an array of different vertebrate
CC   tissue both in vitro and in vivo. OT therapeutics can be used for
CC   treating e.g. neurological conditions deriving from acute, subacute, or
CC   chronic injury to the nervous system, including traumatic, chemical, and
CC   vasal injury and deficits (such as the ischemia resulting from stroke),
CC   together with infectious/inflammatory and tumour induced injury; aging of
CC   the nervous system such as Alzheimer's disease; chronic neurodegenerative
CC   diseases of the nervous system (Parkinson's disease, Huntington's chorea,
CC   amylotrophic lateral sclerosis as well as spinocerebellar degenerations);
CC   and chronic immunological diseases of the nervous system or affecting the
CC   nervous system including multiple sclerosis, for selective ablation of
CC   sensory neurons, e.g. in the treatment of chronic pain syndromes, or in
CC   the treatment of neoplastic or hyperplastic transformations such as may
CC   occur in the central nervous system. The products may also be used in
CC   other therapeutic applications related to their activities. The products
CC   can also be used for detection, diagnosis and drug screening.
XX
SQ   Sequence   889 AA;
```

Query Match 3.6%; Score 120.5; DB 20; Length 889;
Best Local Similarity 20.4%; Pred. No. 0.57;
Matches 144; Conservative 92; Mismatches 250; Indels 219; Gaps 36;

```
Qy     32 RRNQRRRTGRQVSPPDNFTAAAQDLAQSLDANTVTFPANISSMPEFRNWAKG-------- 83

Db     97 rhnakcqlslevfandkeicmikveiqdindnapsfssdqidldisenaapgtrfpltsa 156

Qy     84 -KIDLDSDSIGWYFKYLDPAGATE-SARAVGEYSKIPDGLVKFSVDAEIREIYNEECPVV 141

Db    157 hdpdagenglrtylltrddhglfgldvksrgdgtkfpelviqkald---reqqnhhtlvl 213

Qy    142 TDVSVPLDGRQWSLSIFSFPMFRTAYVAVANVENKEMS--LDVVNDLIEWLNNLADWRYV 199

Db    214 t----aldgge-------pprsatvqinvkvidsndnspvfeapsylvelpentplgtvv 262

Qy    200 VDSEQWINFTN-------DTTY----YVRIRVLRPTYDVPDPTEGLVRTVS--DY----- 241

Db    263 id----lnatdadegpngevlysfssyvpdrv-relfsi-dpktglirvkgnldyeengm 316

Qy    242 ------------------RLTYKAITCEANMPTLVDQGFWIGGQYALTPTSLPQYDVSE 282

Db    317 leidvqardlgpnlipahckvtvklidrndnapsi---gf----------vsvrqgalse 363

Qy    283 AYALHTLTFARPSSAAALAFVW---AG---------LPQGGTAPAGTPAWEQASSGGYLT 330

Db    364 a--------appgtvialvrvtdrdsgkngqlqcrvlggggtgggg-------------- 401

Qy    331 WRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFRQV---AVDEVVVTN 387

Db    402 ----glggpggsvpfkleeny--------dnfyt-------vvtdrpmdretqdeynvti 442

Qy    388 NPAGGGSAP-----TFTVRV------PP---SNAYTNTVFRNTLLETRPSSRRLELPMPP 433

Db    443 vardggspplnstksfaikilddndnpprftkglyvlqvhenni--------------pg 488
```

```
Qy    434 ADFGQTVANNPKIEQ------SLLKETLGCYLVHSKMR-NPVFQLTPASSFGAVSFNNPG 486
Db    489 dslgsvlaqdpdlgqngtvsysilpshigdvsiytyvsvnp----tngaiyalrsfn--- 541

Qy    487 YERTRDLPDYTGIRDS-----FDQNMSTAVAHFRSLSHSCSIVTKTYQGWEG-------- 533
Db    542 feqtkafefkvlakdsgapahlesnatvrvtvldvndnapvivlptlqndtaelqvprna 601

Qy    534 -----VTNVNTPFGQFAHAGLLK------NEEILCLADDLATRLTGVYPATDN------F 576
Db    602 glgylvstvraldsdfgesgrltyeivdgnddhlfeidpssgeirtlhpfwedvtpvvel 661

Qy    577 AAAVSAFAANMLSSVLKSEATSSIIKSVGETAVGAAQSGLAKLPG 621
Db    662 vvkvtdhgkptlsavak-----liirsvs----gslpegvprvng 697


RESULT  11
AAE11961
ID   AAE11961 standard; Protein; 979 AA.
XX
AC   AAE11961;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Streptococcus agalactiae R5 protein.
XX
KW   Cell surface protective antigen; R5 protein; antibacterial; therapeutic;
KW   immunotherapy; group B streptococci infection; prophylactic; vaccine.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200173037-A2.
XX
PD   04-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-EP03618.
XX
PR   30-MAR-2000; 2000EP-0106818.
PR   17-OCT-2000; 2000EP-0122620.
XX
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI   Chhatwal GS,  Erdogan S,  Fagan PK,  Guzman CA,  Talay SR;
XX
DR   WPI; 2001-616506/71.
DR   N-PSDB; AAD17326.
XX
PT   Vaccine for preventing and treating infection by group B streptococci
PT   comprises the cell surface protective antigen of group B streptococci,
PT   R5 protein -
XX
PS   Claim 11; Fig 1; 48pp; English.
XX
CC   The invention relates to new cell surface protective antigen of group B
CC   streptococci, R5 protein and its corresponding DNA molecule. R5 protein
CC   and its DNA are useful for diagnosing an infection with group B
CC   streptococci. A composition comprising R5 protein is useful for
CC   immunotherapy of an infection with group B streptococci and for
CC   prophylactic and therapeutic vaccination against the same infection.
CC   The present sequence is Streptococcus agalactiae R5 protein.
XX
SQ   Sequence   979 AA;


  Query Match             3.6%;  Score 120.5;  DB 22;  Length 979;
  Best Local Similarity   22.2%;  Pred. No. 0.66;
  Matches  107;  Conservative   71;  Mismatches  155;  Indels  149;  Gaps   30;

Qy     19 NVRVS-ANTVTVNGRRNQRRRTGR---------QVSPPDNFTAAAQDLAQSLDANTV--- 65
Db    258 narlsftetvmengkivekfdyargsynstffdgitpgislekassgvnltvtantvdvt 317

Qy     66 ---TFPANISSMPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATESARAVGEYSKIPDGLV 122
Db    318 dkntfnesvvn-psdedfvngp-drtpdav--------pag-tgsirlkgtft------- 359

Qy    123 KFSVDAEIREIYNEECPVVTDVSVPLDGRQWSLSIFSFPMFRTA---YVAVANVENKEMS 179
Db    360 ----dasfk-lyhqavp--------------stafskekyhtgdgyynsianvrptvvk 399

Qy    180 LDVVNDLIEWLNNLADWRYVVDSEQWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVS 239
Db    400 pdsinglnkhasdvidyk-isdnn---disnddllrlsvrlqnprgsv----------vv 445

Qy    240 DYRLTYKAITCEANMPTLVDQGFWIGGQYALTPTSLP--QYDVSEAYA-LHT-LTFARPS 295
Db    446 ny------idtegni---------igteykdttdaipgthyntaessgdlnsdatverps 490

Qy    296 SAAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLTWRHNGTTFPAG------------S 342
Db    491 titkdgkvydlvaenitvpvg-----kvnsdg--tlatngssfnygtdaasaevaegtks 543

Qy    343 VSYVL-----PEGFALERYDPNDGSWTDFASAGDTVTFRQVAVDEVVVTNNPA------- 390
Db    544 viyvisikqeskgnvhary-vilgtetelasa-ktvk-seapideaysdkapatlekdgk 600

Qy    391 ---------GGGSAP----------TFT---VRVPPSNAYTNTVFRNTLLETRPSSRRLE 428
Db    601 lyefvhvrdnkgdapadgkvteqdqtityeyvevpkgrvvvdyviegt--atkpkdtyvd 658

Qy    429 LP 430
Db    659 tp 660


RESULT  12
AAY00217
ID   AAY00217 standard; Protein; 932 AA.
XX
AC   AAY00217;
XX
DT   20-APR-1999  (first entry)
XX
DE   Enterococcus faecalis antigenic polypeptide fragment EF107.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C,  Choi GH,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   N-PSDB; AAX20207.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 9; Page 208; 301pp; English.
XX
CC   The present sequence represents an antigenic polypeptide fragment
```

```
CC   isolated from Enterococcus faecalis. The present invention describes
CC   genes, proteins and antigenic polypeptides isolated from E. faecalis.
CC   The proteins can be used in vaccines for preventing or attenuating an
CC   infection caused by a member of the Enterococcus genus in an animal.
CC   They can also be used for detecting Enterococcus antibodies in a sample.
CC   The nucleotide sequences can be used for detecting Enterococcus nucleic
CC   acids. Products from the present invention can also be used for
CC   screening compounds to identify agonists and antagonists of E. faecalis
CC   protein activity.
XX
SQ   Sequence   932 AA;


  Query Match             3.6%;  Score 120;  DB 20;  Length 932;
  Best Local Similarity   21.0%;  Pred. No. 0.67;
  Matches  159;  Conservative   82;  Mismatches  261;  Indels  254;  Gaps   45;

Qy       2 GDAGVASQRPHNRRGTRNVRVSANTVTVNGRRNQRRRT-------GRQVSPPDNFTAAAQ 54
Db     222 gdvg-gptsshqrrgtag--------pvyyyltkrrvtekfenpaggaipapegyt---q 269

Qy      55 DLAQSLDANTVTFPANISSMPE-----------FRNWAKGKI---DLDSDSIGWY----- 95
Db     270 dkktivtgedftftqe-gtlperytgsdgktylfkgwykgnakpstlettktpsyavtyd 328

Qy      96 ------------------------FKYLDPAG-----------AT--ESARAVGEYSKI 117
Db     329 dnddlhvvyeeavmktytlparealfgyvdeqgnlinpakfklsatmgesdgatgemttf 388

Qy     118 P--DG-------LVKFSVDAEIREIYN--EECPVVT----DVSVPLDGRQWSLSIFSFPM 162
Db     389 ptidgidmpasqlkklaip---qkvytrpddgtivtygpqevsveipkyyqtisisp--- 442

Qy     163 FRTAYVA------VAN-----VENKEMSLDVVNDLIEW----LNNLADWRYVVDSEQWI- 206
Db     443 -ttaytgdktkypvpnevrrgienpd---nivsslvgxxaynltqksatrytarrsywxw 498

Qy     207 ------------------NF---TNDTTYYVRIRVLRPTYDVPDP-----------TEGL 234
Db     499 gptktlysmsiysgtaggnynlstpdgtiyyylenrrvtehfvdesgakitpptgftqgn 558

Qy     235 VRTVSDYRLTY---KAI--TCEANMPTLVDQGFWIGGQY-----------ALTPTSLPQY 278
Db     559 qlvvdsenyvytvakalpkiyqagektyifqg-wfkgktkpatlkttttpsftptfnded 617

Qy     279 DVSEAY------ALHTLTFARP--SSAAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLT 330
Db     618 dmtavyqeaiptaeltltgavdiiengatmdyweallkntgeaplttikikpta-----t 672

Qy     331 WRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFRQVAVDEVVVTNNPA 390
Db     673 w---aagigapntifvqgtgqntkafpvtkeqwt--tgagvsitldqpl---------pa 718

Qy     391 GG-------GSAPT------FTVRVPPSNAYTNTVFRNTLLETRPSSRRLELP------- 430
Db     719 ggqlkmnllgtavtgnpgqvltadvevtgnfgsltakdt-vrikdldqeitspdgdgfis 777

Qy     431 MPPADFGQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQL-TPASSFGAVSFNNPGYER 489
Db     778 tptfdfgklaisgsk-qqyglkkaadyy--gngtrnpylrlntsqanwsltaqlsqpksa 834

Qy     490 TRDLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFGQFAHAGL 549
Db     835 tdslptttrll------lgtaaa------------asftdynqptetrtplg------- 868

Qy     550 LKNEEILCLADDLAT------RLTG--VYPATDNFA 577
Db     869 -ktstvtltadntatavvanqqftgsdvyqldftfa 903


RESULT  13
AAY00216
ID   AAY00216 standard; Protein; 969 AA.
XX
AC   AAY00216;
XX
DT   20-APR-1999  (first entry)
XX
DE   Enterococcus faecalis protein EF107.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08959.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C,  Choi GH,  Hromockyj A,  Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   N-PSDB; AAX20206.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 9; Page 206-207; 301pp; English.
XX
CC   The present sequence represents a protein isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
XX
SQ   Sequence   969 AA;


  Query Match             3.6%;  Score 120;  DB 20;  Length 969;
  Best Local Similarity   21.0%;  Pred. No. 0.71;
  Matches  159;  Conservative   82;  Mismatches  261;  Indels  254;  Gaps   45;

Qy       2 GDAGVASQRPHNRRGTRNVRVSANTVTVNGRRNQRRRT-------GRQVSPPDNFTAAAQ 54
Db     259 gdvg-gptsshqrrgtag--------pvyyyltkrrvtekfenpaggaipapegyt---q 306

Qy      55 DLAQSLDANTVTFPANISSMPE-----------FRNWAKGKI---DLDSDSIGWY----- 95
Db     307 dkktivtgedftftqe-gtlperytgsdgktylfkgwykgnakpstlettktpsyavtyd 365

Qy      96 ------------------------FKYLDPAG-----------AT--ESARAVGEYSKI 117
Db     366 dnddlhvvyeeavmktytlparealfgyvdeqgnlinpakfklsatmgesdgatgemttf 425

Qy     118 P--DG-------LVKFSVDAEIREIYN--EECPVVT----DVSVPLDGRQWSLSIFSFPM 162
Db     426 ptidgidmpasqlkklaip---qkvytrpddgtivtygpqevsveipkyyqtisisp--- 479

Qy     163 FRTAYVA------VAN-----VENKEMSLDVVNDLIEW----LNNLADWRYVVDSEQWI- 206
Db     480 -ttaytgdktkypvpnevrrgienpd---nivsslvgxxaynltqksatrytarrsywxw 535
```

```
Qy    207 ------------------NF---TNDTTYYVRIRVLRPTYDVPDP-----------TEGL 234
                           |:   | | | |  :   | |    |            |:|
Db    536 gptktlysmsiysgtaggnynlstpdgtiyyylenrrvtehfvdesgakitpptgftqgn 595

Qy    235 VRTVSDYRLTY---KAI--TCEANMPTLVDQGFWIGGQY-----------ALTPTSLPQY 278
             |      |   ||:    :|   | : || |  |:            : |||   :
Db    596 qlvvdsenyvytvakalpkiyqagektyifqg-wfkgktkpatlkttttpsftptfnded 654

Qy    279 DVSEAY------ALHTLTFARP--SSAAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLT 330
            |::  |      |  ||| |     : | : :  | |   | ||  |   :  :     |
Db    655 dmtavyqeaiptaeltltgavdiiengatmdyweallkntgeaplttikikpta-----t 709

Qy    331 WRHNGTTFPAGSVSYVLPEGFALERYDPNDGSWTDFASAGDTVTFRQVAVDEVVVTNNPA 390
            |       | :  :|   |   : :      ||    || ::|  |           ||
Db    710 w---aagigapntifvqgtgqntkafpvtkeqwt--tgagvsitldqpl---------pa 755

Qy    391 GG-------GSAPT------FTVRVPPSNAYTNTVFRNTLLETRPSSRRLELP------- 430
            ||       |:| |       |  |  :  : :   ::| :  :   : :  |
Db    756 ggqlkmnllgtavtgnpgqvltadvevtgnfgsltakdt-vrikdldqeitspdgdgfis 814

Qy    431 MPPADFGQTVANNPKIEQSLLKETLGCYLVHSKMRNPVFQL-TPASSFGAVSFNNPGYER 489
             |  |||:   :  | :|  ||:    |   :  |||  :| |  :::   :  :
Db    815 tptfdfgklaisgsk-qqyglkkaadyy--gngtrnpylrlntsqanwsltaqlsqpksa 871

Qy    490 TRDLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIVTKTYQGWEGVTNVNTPFGQFAHAGL 549
            |  ||  | :       : || |              ::  :   |   || |
Db    872 tdslptttrll------lgtaaa------------asftdynqptetrtplg------- 905

Qy    550 LKNEEILCLADDLAT------RLTG--VYPATDNFA 577
            |   :   ||: ||      : ||  ||     ||
Db    906 -ktstvtltadntatavvanqqftgsdvyqldftfa 940


RESULT  14
AAP50073
ID   AAP50073 standard; Protein; 907 AA.
XX
AC   AAP50073;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Epstein-Barr virus (EBV) outer surface protein.
XX
KW   Epstein-Barr virus; antigen; vaccine.
XX
OS   Epstein-Barr virus.
XX
PN   EP151079-A.
XX
PD   07-AUG-1985.
XX
PF   28-JAN-1985;   85EP-0400141.
XX
PR   23-JUL-1984;   84US-0633558.
PR   30-JAN-1984;   84US-0575352.
XX
PA   (UYCH-) UNIV OF CHICAGO.
XX
PI   Kleff E,  Tanner J,  Hummel M,  Belsel C;
XX
DR   WPI; 1985-191978/32.
DR   N-PSDB; AAN50114.
XX
PT   New fragment of Epstein-Barr Virus DNA - useful in vector to
PT   express polypeptide for use in prepn. of vaccine against the
PT   virus and for use in diagnosis.
XX
PS   Claim 2; Page 23-25; 26pp; English.
XX
CC   The sequence encodes an outer surface viral protein of EBV, used
CC   to generate antibodies reacting with the surface proteins of
CC   EBV-infected cells, and in the preparation of a vaccine against EBV.
XX
SQ   Sequence   907 AA;


  Query Match             3.5%;  Score 119.5;  DB 6;  Length 907;
  Best Local Similarity   19.5%;  Pred. No. 0.71;
  Matches  136;  Conservative   67;  Mismatches  274;  Indels  219;  Gaps   27;

Qy     48 NFTAAAQDLAQSLDANTVTFPANISSMPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATES 107
          | ||  :  || ||   || | :: :  :  |::     : :: :|     :|     |
Db    169 nitavvr--aqgld---vtlplslptsaqdsnfs-----vktemlg---neid----iec 211

Qy    108 ARAVGEYSKIPDGLVKFSVDAEIRE-------IYNEECPVVTDVSVPLDGRQWSLSIFSF 160
             || |::  |  ||::     |       |     || |   :|  |  :|| :
Db    212 imedgeisqvlpgdnkfnitcsgyeshvpsggiltstspvat--pipgtgyayslrltpr 269

Qy    161 PMFR------TAYVAVANVENKEMSLD-------VVNDLI----EWLNNLADWRYVVDSE 203
          |: |        ||  :    |    |        | :| |    :   |  |  || |:
Db    270 pvsrflgnnsilyvfysgngpkasggdyciqsnivfsdeipasqdmptnttdityvgdn- 328

Qy    204 QWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCEANMPTLVDQGFW 263
                              || ||  |                :|| | :    ||
Db    329 -------------------atysvpmvtse----------------danspnvtvtafw 352

Qy    264 --------------------------IGGQYALTPT----------------------- 273
                                       | | :|   |
Db    353 awpnntetdfkckwtltsgtpsgcenisgafasnrtfditvsglgtapktliitrtatna 412

Qy    274 ---------SLPQYDVSEAYALHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQAS 324
                   |      : :  |:|  || |::          |||     |    |   ||
Db    413 ttthkvifskapestttsptlnttgfadpntt-------tglpssthvptnlta--pas 463

Qy    325 SGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGSW-----------TDFASAGDTV 373
          :|  ::     :  |||: |   |          |:   |           |   |   |
Db    464 tgptvstadvtsptpagttsgasp-------vtpspspwdngteskapdmtsstspvttp 516

Qy    374 TFRQVAVDEVVVT-------------------NNPAGGGSAPTFTVRVPPSNAYTNTVFR 414
          |    :    | |                    :|  | ::||  |  |  || : |: :
Db    517 tpnatsptpavttptpnatsptpavttptpnatsptlgktsptsavttptpnatsptlgk 576

Qy    415 N--------------TLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCY 458
                          || :| |:|  :  | | |  | ||      : :    |||
Db    577 tsptsavttptpnatsptlgktspts-avttptpnat-gptvgetsp-qanatnhtlg-- 631

Qy    459 LVHSKMRNPVFQLTPASSFGAVSFNNPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLS 518
                ||    | ::  ||:     :  |        :| | :    :      | |
Db    632 ---gtsptpvvtsqpknatsavttgq--hnitssstssmslrpssnpetlspstsdnsts 686

Qy    519 HSCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEILCLADDLATRLTGVYPATDNFAA 578
          |   : :    | | :| | ||     |        :   : :     |       | :
Db    687 hmplltsahptggenitqv-tpasisth-------hvstsspeprpgttsqasgpgnsst 738

Qy    579 AVSAFAANMLSSVLKSEATSSIIKSVGETAVGAAQS 614
          :      |:        |||    |  :|||    |
Db    739 stkpgevnvtkgtppqnatspqapsgqktavptvts 774


RESULT  15
AAY68009
ID   AAY68009 standard; protein; 907 AA.
XX
AC   AAY68009;
XX
DT   11-APR-2000  (first entry)
XX
DE   Epstein-Barr virus major outer envelope glycoprotein Gp350.
XX
KW   Epstein-Barr virus; major outer envelope glycoprotein; Gp350; Gp220;
KW   B cell activation; immunoglobulin; secretion; EBV; Gp350/220; antigen;
KW   vaccine; adjuvant; fusion protein; antimicrobial; antiviral; CR2; CD21;
KW   anti-allergic; antitumour; immunostimulatory; infection; tumour;
KW   complement receptor 2.
XX
OS   Epstein-barr virus.
```

XX
PN WO9964603-A2.
XX
PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99WO-US13113.
XX
PR 12-JUN-1998; 98US-0089158.
XX
PA (JACK-) JACKSON FOUND ADVANCEMENT MILITARY MED.
PA (MOND/) MOND J J.
PA (LEES/) LEES A.
XX
PI Mond JJ, Lees A;
XX
DR WPI; 2000-116546/10.
XX
PT Adjuvant containing an Epstein-Barr virus sequence that binds to
PT complement receptor 2, for protecting against, e.g. infection or tumors
PT -
XX
PS Disclosure; Fig 1A; 43pp; English.
XX
CC The present invention describes an adjuvant (A) comprising at least one
CC EBV (Epstein-Barr virus) Gp350/220 sequence (I) that binds to CR2
CC (complement receptor 2; CD21). Vaccines containing (A) and at least one
CC other antigenic epitope (also similar immunogens and nucleic acid
CC vectors that express (I)) are used to elicit an antibody response to
CC both (I) (for protection against EBV) and the second epitope, which may
CC be associated with some other infectious agent, allergen, tumour antigen
CC or generally any condition that requires immune stimulation. The vaccines
CC and immunogens may also be used to treat B cells in vitro. (A) increases
CC the immunogenicity of a co-administered antigen. By combining (I), a
CC non-specific B cell stimulator, with another antigen, the stimulatory
CC effect is targeted to antigen-specific B cells, i.e. B cells that
CC secrete immunoglobulins specific for the antigen are stimulated
CC simultaneously through the antigen receptor and through (I)-mediated
CC crosslinking of CR2. Constructs of (I) and second antigen are relatively
CC small, so should be cleared relatively quickly (reducing immunogenicity)
CC and many copies of second antigen can be included to increase
CC antigenicity. The need for expensive anti-CR antibodies and potentially
CC inflammatory C3 components is avoided and folding of the CR2-binding
CC domain in (I)-antigen fusion proteins is unlikely to be disrupted. The
CC present sequence represents EBV major outer envelope glycoprotein Gp350.
XX
SQ Sequence 907 AA;

Query Match 3.5%; Score 119.5; DB 21; Length 907;
Best Local Similarity 19.5%; Pred. No. 0.71;
Matches 136; Conservative 67; Mismatches 274; Indels 219; Gaps 27;

```
Qy     48 NFTAAAQDLAQSLDANTVTFPANISSMPEFRNWAKGKIDLDSDSIGWYFKYLDPAGATES 107
Db    169 nitavvr--aqgld---vtlplslptsaqdsnfs-----vktemlg---neid----iec 211

Qy    108 ARAVGEYSKIPDGLVKFSVDAEIRE-------IYNEECPVVTDVSVPLDGRQWSLSIFSF 160
Db    212 imedgeisqvlpgdnkfnitcsgyeshvpsggiltstspvat--pipgtgyayslrltpr 269

Qy    161 PMFR------TAYVAVANVENKEMSLD-------VVNDLI----EWLNNLADWRYVVDSE 203
Db    270 pvsrflgnnsilyvfysgngpkasggdyciqsnivfsdeipasqdmptnttdityvgdn- 328

Qy    204 QWINFTNDTTYYVRIRVLRPTYDVPDPTEGLVRTVSDYRLTYKAITCEANMPTLVDQGFW 263
Db    329 ----------------------atysvpmvtse----------------danspnvtvtafw 352

Qy    264 --------------------------IGGQYALTPT------------------------ 273
Db    353 awpnntetdfkckwtltsgtpsgcenisgafasnrtfditvsglgtapktliitrtatna 412

Qy    274 ---------SLPQYDVSEAYALHTLTFARPSSAAALAFVWAGLPQGGTAPAGTPAWEQAS 324
Db    413 tttthkvifskapestttsptlnttgfadpntt-------tglpssthvptnlta--pas 463

Qy    325 SGGYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGSW----------TDFASAGDTV 373
Db    464 tgptvstadvtsptpagttsgasp-------vtpspspwdngteskapdmtsstspvttp 516

Qy    374 TFRQVAVDEVVVT------------------NNPAGGGSAPTFTVRVPPSNAYTNTVFR 414
Db    517 tpnatsptpavttptpnatsptpavttptpnatsptlgktsptsavttptpnatsptlgk 576

Qy    415 N---------------TLLETRPSSRRLELPMPPADFGQTVANNPKIEQSLLKETLGCY 458
Db    577 tsptsavttptpnatsptlgktspts-avttptpnat-gptvgetsp-qanatnhtlg-- 631

Qy    459 LVHSKMRNPVFQLTPASSFGAVSFNNPGYERTRDLPDYTGIRDSFDQNMSTAVAHFRSLS 518
Db    632 ---gtsptpvvtsqpknatsavttgq--hnitssstssmslrpssnpetlspstsdnsts 686

Qy    519 HSCSIVTKTYQGWEGVTNVNTPFGQFAHAGLLKNEEILCLADDLATRLTGVYPATDNFAA 578
Db    687 hmplltsahptggenitqv-tpasisth------hvstsspeprpgttsqasgpgnsst 738

Qy    579 AVSAFAANMLSSVLKSEATSSIIKSVGETAVGAAQS 614
Db    739 stkpgevnvtkgtppgnatspqapsgqktavptvts 774
```

Search completed: May 29, 2002, 11:39:37
Job time: 50 sec